

Tózsér János

Mit jelent személynek lenni?

BEVEZETÉS

Mi emberek mennyiben vagyunk *többek* az állatoknál, amelyek hozzánk hasonlóan ugyancsak lélegeznek, táplálkoznak, észlelik környezetüket, és képesek fájdalmat érezni? Illetve: mennyiben vagyunk *többek* annál a szuperszámítógépnél, amelyet minden kétséget kizáróan intelligensnek mondunk, hiszen Garri Kaszparovot is képes volt legyőzni sakkban? A filozófiában rendszerint a „személy” fogalma az, amelynek segítségével *kitüntetjük* magunkat mind az állatokkal, mind pedig a számítógépekkel szemben: mi emberek *személyek* vagyunk, az állatok és a számítógépek ezzel szemben nem személyek.

Ebben a tanulmányban azt a problémát fogom vizsgálni, hogy *mit jelent* személynek lenni. Mielőtt azonban nekilátnék e probléma tárgyalásának, szeretném pontosan körvonalazni azt, amivel foglalkozni fogok.

Szokás megkülönböztetni a „személy” fogalmának kétféle, *metafizikai* és *morális* jelentését. „Személy” alatt *metafizikai* értelemben olyan lényeket értünk, amelyek érzésekkel, érzetekkel, gondolatokkal, vágyakkal stb. rendelkeznek, amelyek inkább események elindítói, iniciálói, mintsem környezetükre pusztán ösztönösen avagy mechanikusan reagáló alakok, és amelyekről inkább azt mondjuk, hogy képesek a külvilágról különféle ideák kialakítására, mintsem azt, hogy pusztá előforduló dolgok a világban. *Morális* értelemben véve személynek lenni mást jelent. Morális értelemben személynek lenni ugyanis annyi, mint joggal elvárni bizonyos privilégiumokat mindazoktól, akiknek cselekedetei befolyásolják jólétünket. Morális értelemben tehát azokat a lényeket nevezzük személyeknek, akiknek az érdekeit figyelembe *kell* vennünk, akiknek jogokat és kötelességeket *kell* tulajdonítanunk. A metafizikai értelemben vett személy-fogalom tehát *deskriptív*, a világban lévő dolgok egy osztályát jelölő fogalom, a morális értelemben vett „személy” ezzel szemben *értékelő* fogalom, amely a dolgok egy osztályát bizonyos privilégiumokkal ruházza fel.

A kétfajta személy-fogalom számos filozófiai problémát vet fel. A legelemibb ezek közül az, hogy *milyen viszony* van a metafizikai értelemben vett és a morális értelemben vett személy-fogalom között. Vajon azonosak-e, egybeesnek-e, vagy inkább arról van-e szó, hogy a metafizikai értelemben vett „személy” csak szükséges feltétele a morális értelemben vett „személy” fogalmának? Vagy, ugyanez a probléma más megfogalmazásban: Rawls az *A Theory of Justice*-ban azt vizsgálja-e, hogy az eredeti állapotban levő metafizikai személyek *hogyan válnak* morális személyekké, vagy azt, hogy a metafizikai személyek miért mondhatók *együttal* morális értelemben vett személyeknek.¹

A „személy” mind metafizikai, mind pedig morális fogalma kapcsán felmerül az előzőn felül a következő probléma: koextenzív-e a „személy” és az „ember” fogalma?

¹ Ezzel a kérdéssel részletesen foglalkozik SAPONTZIS 1981, és konkrétan Rawls vonatkozásában RORTY 1973.

De facto csak az embereket nevezzük személynek, azonban elképzelhetünk olyan idegen (földön kívüli és természetfeletti) lényeket, amelyeket ugyancsak személyeknek neveznénk, valamint ismerünk olyan eseteket (például súlyos értelmi visszamaradottság), ahol az illető embert metafizikai értelemben nem nevezzük személynek.

Ebben a tanulmányban nem ezekkel a kérdésekkel fogok foglalkozni, hanem azzal, hogy mit jelent metafizikai értelemben véve személynek lenni. Minek kell teljesülnie ahhoz, hogy egy lényről (vagy rendszerről) jogosan azt mondhassuk, hogy az személy?

Ha egy pillantást vetünk a *conditions of personhood* probléma irodalmára, azt találjuk, hogy a különböző szerzők megoldási javaslataikban a „személy” fogalmának igen nagy mennyiségű szükséges feltételét halmozták fel. A következő listát kínálom most:

Egy adott *S* rendszert személynek nevezünk abban az esetben,

- ha megtestesült (*embodied*);
- ha él;
- ha rendelkezik nyelvi képességgel;
- ha lehetőségünkben áll mentális állapotokat tulajdonítani neki;
- ha ugyancsak képes mentális állapotokat tulajdonítani más rendszereknek;
- ha rendelkezik egy sajátos, csak a személyekre jellemző tudattal (*consciousness*);
- ha rendelkezik szabad akarattal;
- ha fennmaradása során mindvégig ugyanazzal a testtel rendelkezik;
- ha helyes emlékképei vannak önmagáról;
- ha önmaga alatt az időben mindig ugyanazt ismeri meg stb.

Természetesen felvethető az a kérdés, hogy lehetőségünkben áll-e megadni a „személynek lenni” szükséges feltételeit úgy, hogy együttesen a „személynek lenni” *elég-séges* feltételeit is megadják. Ezzel a problémával ezen a helyen megint csak nem foglalkozom. Ehelyett az iménti lista három – véleményem szerint legfontosabb – feltételét tárgyalom: azt jelesül, hogy 1. a személy fogalma milyen viszonyban áll a mentális állapotok tulajdonításának problémájával; 2. mit jelent csak a személyekre jellemző öntudattal rendelkezni; és 3. mit jelent szabad akarattal rendelkezni. E három feltétellel szervesen összekapcsolódik a „személy” fogalmának három meghatározása: Strawsoné, Nagelé, valamint Frankfurté. Ebben a tanulmányban amellet érvelek majd, hogy mindhárom felfogás elhibázott, és hogy valójában nem tudjuk, mit jelent személynek lenni.

A HÁROM JAVASLAT

1. Strawson „Persons” című tanulmányában a következőképpen definiálja a „személy” fogalmát:

„A »személy« fogalmán egy olyan típusú entitás fogalmát értem, hogy a tudatos állapotokat tulajdonító predikátumok és testi jellemzőket tulajdonító predikátumok... egyformán alkalmazhatóak ennek a típusnak egy példányára.” (STRAWSON 1962, 135.)

Strawson e definícióban azt állítja, hogy a „személy” fogalma nem más, mint fizikai és mentális predikátumok *logikai szubjektuma*, más szóval azt, hogy a mentális predikátumokat *pontosan ugyanannak a dolognak* tulajdonítjuk, mint a fizikai predikátumokat. Ez az álláspont egyáltalán nem triviális, hiszen léteznek ettől gyökeresen eltérő felfogások is. Ilyen például a karteziánus álláspont, amely szerint a mentális predikátu-

mok a különböző elmékre, a fizikai predikátumok pedig a különböző testekre vonatkoznak. És ugyancsak ide sorolható a Strawson által „no-ownership”-nek vagy „no-subject”-nek nevezett elmélet, amely szerint valójában mi egyáltalán nem tulajdonítunk mentális avagy tudatos állapotokat.² Strawson tehát e két állásponttal szemben azt állítja, hogy egyrészt igenis *tulajdonítunk* mentális állapotokat, másrészt ezeket a mentális predikátumokat *ugyanannak a dolognak* (to the very same thing) tulajdonítjuk, mint a fizikai predikátumokat.

Továbbmenve: amennyiben azt állítjuk, hogy a „személy” fogalma nem más, mint aminek mentális és fizikai predikátumokat tulajdonítunk, akkor ezzel azt állítjuk, hogy a „személy” fogalma egy *logikailag primitív, tovább nem analizálható* fogalom. Mit jelent ebben az összefüggésben az, hogy tovább nem analizálható? Azt, hogy a személyre nem úgy kell gondolnunk, „mint kétféle alany egyfajta összetételére, egyrészt mint a tapasztalatok alanyára (tudatos egóra), másrészt mint a testi tulajdonságok alanyára” (STRAWSON 1962, 136). Magyarán a „személy” fogalmát nem úgy kell megragadnunk, mint két különböző szubjektumból (testből és tudatból) álló rendszert. Strawson elképzelései szerint tehát a „személy” fogalmához nem a tudat filozófiai analízise útján jutunk: Hume-nak teljesen igaza volt, amikor azt állította, hogy ha introspektíve magunkba pillantunk, nem találunk semmiféle Én-szubsztanciát, hanem csak percepciókat, illetve percepciók kötegét. Strawson álláspontjához azonban nem Hume, hanem Kant elemzése áll igazán közel. Kant volt ugyanis az, aki azt vetette mind a racionalisták, mind pedig az empiristák szemére, hogy valamennyien úgy értelmezik az Ént, mint egy *kvázi objektumot*. Kant szerint e két álláspontnak végső soron az a közös hibája, hogy azt képzelik, ahogy az empirikus tárgyakat a külső megfigyelés adataiból, úgy az Ént, azaz a személyes szubjektumot a belső megfigyelés adataiból konstruáljuk. E helyett – Strawsonnak igen tetsző módon – Kant azt állította, hogy a személyes szubjektum nem valamilyen belső észlelés tárgya, sem belső észlelések összessége vagy együttese, hanem az Én, vagyis az, ami a tudat egységét, identitását adja, *egy tartalmilag üres, formális képzet*, a megismerés pusztá feltétele.

2. Biztosra vehetjük azt, hogy mi, személyek *tudatos* lények vagyunk, és bennünket más lényektől e sajátos *tudatosságunk*, vagy e sajátos *tudatunk* különböztet meg. Van egy módja annak, *ahogyan* tudatosak vagyunk, és ezzel egyetlen más osztályba tartozó dolog, sem állat, sem pedig semmiféle számítógép nem rendelkezik. Ez a *csak ránk*, emberekre jellemző tudatosság vagy tudat (consciousness) kimondhatatlan, elemezhetetlen, primitív fogalom. Nem áll módunkban ugyanis e sajátos minőséghez hozzáférni másképp, mint *belülről* (from the inside); csak megtapasztalva, átélve ismerhető meg a tudat e minősége.

Nagel a „What is it Like to be a Bat?” című tanulmányában a denevér példáján keresztül mutatja be a tudat e minőségét:

„Az a tény, hogy egy organizmusnak egyáltalán vannak tudatos tapasztalatai, lényegében azt jelenti, hogy valamilyen ennek az organizmusnak lenni. Alapvetően egy or-

² Strawson szerint Wittgenstein és Schlick képviselte az Én „no-ownership” vagy „no-subject” elméletét. Wittgenstein például az „I have a toothache” mondat kapcsán – ellentétben az „I have a bad tooth” mondattal, ahol is az „I” szó rám mint fizikai testre referál – azt állította, hogy ebben a mondatban az „I” *semmiféle tulajdonost* nem jelöl. Illetve, ha vesszük az „I think” vagy „ich denke” kijelentést, akkor valójában úgy kellene fogalmaznunk, hogy „there is a thought”, vagy hogy „es denkt”. Schlick esetében éppen ilyen világos a helyzet. Schlick a „Meaning and Verification” című tanulmányában ekképp fogalmaz: „Beláttuk, hogy hacsak nem azt mondjuk, hogy érzetadataink hordozója vagy viselője a testünk – noha ez abszolút félrevezető volna –, azt kell mondanunk, hogy érzetadatainknak nincsenek hordozói, avagy viselői.” (SCHLICK 1949.)

ganizmus akkor és csak akkor rendelkezik tudatos mentális állapotokkal, ha van olyan valami, mint annak az organizmusnak *lenni* – ha az organizmus *számára* a világ valamilyen.” (NAGEL 1979, 166.)³

Nagel cikkének a fő mondanivalója az, hogy még ha feltételezzük is, hogy pontosan ismerjük a denevér érzékelésének minden releváns neurofiziológiai, pszichológiai, élettani stb. vonatkozását, akkor is az a mód, *ahogyan* a denevér érzékel (ahogy tehát a denevér *conscious*), szükségképpen ismeretlen marad. Nem tudjuk, *milyen is az*, a denevér érzékleteivel rendelkezni. Hiányoznak eszközeink arra, hogy megtudjuk, *fenomenológiailag* milyen a denevér érzékleteivel rendelkezni, holott e számunkra megragadhatatlan minőségek teszik a denevér érzékelését *azzá*, mint ami. Egyszóval, a denevér mentális állapotainak meghatározott szubjektív karaktere *túl van* megismerőképességünk fogalmainak határán.

Nagel ebben a tanulmányában amellet érvel, hogy semmit nem segít az, ha elképzeljük, milyen denevérek lenni:

„Nem segít, ha elképzeljük azt, hogy kezünkön repülőhártya van, amely képessé tesz bennünket a sötétség beálltakor, valamint szürkületben körbe-körbe repdesni, és azt, hogy mindeközben szánkkal rovarokat fogdosunk. Elképzelhetjük, hogy gyenge látóképességgel rendelkezünk, és hogy környezetünket magas frekvenciatarományokhoz tartozó jelek segítségével érzékeljük, és hogy napközben fejjel lefelé lógunk. Amikor mindezt elképzeljük, akkor ugyanis ez azt mutatja, hogy milyen volna *nekem* úgy viselkednem, mint ahogy a denevér viselkedik. A kérdés azonban nem ez. Jómagam ugyanis azt szeretném tudni, hogy milyen a *denevérek*, hogy milyen is az, denevérek lenni.” (NAGEL 1979, 168.)

A „tudat” vagy „tudatosság” fogalmának nageli értelmezésével kapcsolatban feltétlenül el kell oszlatnom egy nagyon gyakori félreértést. Mentális állapotaink szubjektív karakterét sokan összetévesztik mentális állapotaink állítólagos *privát* jellegével, azzal a nézettel tehát, miszerint minden ember *csak a saját maga* tudattartalmaihoz fér hozzá közvetlenül. Nagel nem ezt állítja. Amit Nagel hangsúlyozni akar, az élményeink *bensőségessége*. Mentális állapotaink, érzékleteink, érzeteink, vágyaink, gondolataink stb. lényegi természetüknél fogva *élmény-perspektívánkkal* vannak összekapcsolódva: mentális állapotaink minőségei (a qualiák!) saját perspektívánkhoz kötöttek. Úgy is fogalmazhatok, hogy ezekre a qualiákra maradéktalanul igaz az *esse est percipi* (vagy *esse est experiri*) elv, ugyanis egyrészt ezek a minőségek *egyedül számunkra* léteznek, másrészt *csak azáltal* léteznek, hogy *számunkra* léteznek. Azonban ez *nem* egyes szám első személyben értendő. Míg ugyanis – mondja Nagel – nem áll módomban elképzelni egy denevér érzékleteit, addig igenis módomban áll elképzelni egy *embertársam* érzéseit, pontosan tudhatom azt, hogy *milyen is neki* most.

Nagel „consciousness” fogalmára támaszkodva Denny a következő személy-meghatározással állt elő:

„Mi, személyek (emberek) sajátos minőségű tudattal rendelkező lények vagyunk. Érzékleteink, érzeteink, gondolataink, vágyaink minősége csak belülről ismerhető meg: egyedül *mi emberek* tudjuk azt, hogy *milyen is* érzékleteinkkel, érzeteinkkel, gondolatainkkal stb. rendelkezni. Egyszóval, személynek lenni nem más, mint egy ki-

³ Az angol eredetinek a második mondata a következőképpen hangzik: „Fundamentally an organism has conscious mental states if and only if there is something that it is like to be that organism – something it is like for the organism.”

mondhatatlan és kívülről (más faj nézőpontjából) nem megismerhető tudattal, *Nagel-tudattal* rendelkezni.” (DENNY 1986, 211.)

3. Frankfurt a „Freedom of the Will and the Concept of a Person” című tanulmányában az eddigiektől egészen eltérő módon határozza meg a „személy” fogalmát:

„Azon a nézeten vagyok, hogy a lényeges különbség személyek és nem személyek között a személy akaratának struktúrájában található. Az emberek nem az egyetlen olyan lények, amelyeknek kívánságaik és motívumaik vannak ..., nem ebben különböznek más fajok tagjaitól. (...) Úgy tűnik, hogy az ember különleges sajátossága az, hogy képes önmagában – ahogy én nevezem – »másodfokú vágyak« kialakítására. A kívánságok és választások mellett, hogy ezt vagy azt tegye, az embernek lehetőségében áll ezen felül az, hogy arra vágyjon, meghatározott vágyai vagy motívumai legyenek vagy ne legyenek. (...) Sok állat képes »elsőfokú vágyakra«, de az emberen kívül egyetlen állatnak sincsen képessége arra, hogy reflektáló önértékelése legyen, amit másodfokú vágyai megalkotása fejez ki.” (FRANKFURT 1971, 6.)

Különbséget kell tehát tennünk – mondja Frankfurt – *elsőfokú* és *másodfokú* vágyak és kívánságok között. Elsőfokú vágyaink mindig valamilyen konkrét cselekvés megtételére vagy nem megtételére irányulnak, másodfokú vágyaink azonban vágyainkra és kívánságainkra. Az embernek mint személynek lehetőségében áll tehát *azon felül*, hogy ezt és ezt a cselekedet meg akarja tenni, az *is*, hogy azt akarja, hogy ilyen és ilyen *akarattal* rendelkezzen. Sőt, léteznek olyan esetek, amikor az akarat struktúrája ennél is bonyolultabb struktúrát mutat. Képzeljünk el egy orvost, aki drogfüggőkkel foglalkozik. Ez az orvos gondolhatja azt, hogy betegeit lényegesen eredményesebben tudná kezelni abban az esetben, ha pontosan tudná (érezné) azt, hogy milyen is az, ellenállhatatlan kényszert érezni a drog bevitelére, anélkül azonban, hogy azt akarná, hogy ez a kívánsága hatékony legyen (FRANKFURT 1971, 8–10).

Frankfurt álláspontja szerint csakis másodfokú vágyai miatt mondjuk az embert *szabad* lénynek. Csak azért lehetséges az, hogy egyszer örüljünk szabadságunknak, másszor pedig hiányoljuk azt, mert képesek vagyunk másodfokú kívánságok kialakítására. Egyszóval tehát egy személy akaratát abban az esetben nevezzük szabadnak, ha a személy abban a tekintetben szabad, vagyis autonóm, hogy épp milyen vágyakkal és kívánságokkal rendelkezik. Más szóval: egy adott személy akarata akkor szabad, ha *lehetséges* volna, hogy *más* akarattal rendelkezzen, mint amivel ténylegesen rendelkezik, ha *akarhatna* mást is, mint amit ténylegesen akar, és egyedül *rajta múlik* az, hogy éppen milyen akarattal rendelkezik.

Strawson kapcsán említést tettem Kantról, talán ez Frankfurt kapcsán sem egészen indokolatlan. Kant a *Gyakorlati ész kritikájában* a „személyes szubjektum” fogalmát az „autonómia” fogalmával kapcsolta össze: a „személyes szubjektum” fogalma egy autonóm módon cselekedni képes lényt előfeltételez. Kant az „autonómia” fogalmával nem azt kívánta hangsúlyozni, hogy a személy képes egy célt kijelölni, és azt ilyen és ilyen módon elérni vagy megvalósítani, hanem azt, hogy a személy képes (vagy képesnek kell lennie!) arra, hogy *saját kívánságait, céljait, értékrendjét meghatározza*, illetve kiválassza. Ebben a tekintetben az „öntételezés” fogalma a döntő: a cselekvő személy esetében fel kell tételeznünk azt (vagyis a „cselekvő személy” fogalmának *transzcendentális lehetőségfeltétele* az), hogy annak lehetőségében áll *autonóm* módon – szemben a természeti hajlandóságaival – meghatározni a maga elsőfokú akaratát.

Íme, ezek voltak a „személy” fogalmának – véleményem szerint – legfontosabb, már-már klasszikusnak számító meghatározásai. E három nézet mindegyike úgy állít

ja be önmagát, mint ami a „személy” fogalmát kellő alapossággal ragadta meg, és tette plauzibilissé. E tanulmány hátralévő részében ezeknek az álláspontoknak nyújtom részletes bírálatát.

MIT JELENT MENTÁLIS PREDIKÁTUMOT TULAJDONÍTANI?

Strawson szerint – mint láttuk – a „mit jelent személynek lenni?” kérdésre az a helyes felelet, hogy „»személynek lenni« nem más, mint mentális és fizikai predikátumok logikai alanyának lenni”.

Strawson álláspontjával szemben a legelemibb ellenvetés egyszerűen az, hogy ezzel a definícióval Strawson a „személy” fogalmának ugyan egy szükséges feltételét mondta ki, ez azonban távolról sem tűnik elégségesnek. Ha vesszük például a „fájdalmat érez”, vagy az „inni *akar*” mentális predikátumokat, akkor ezeket egyes *állatoknak is* tulajdoníthatjuk, vagyis olyan lényeknek, melyeket éppenséggel nem nevezünk személyeknek. Strawson tehát a következő dilemma elé kerül: *vagy* azt kényszerül állítani, hogy mentális predikátumokat állatoknak *is* tulajdoníthatunk, ebben az esetben azonban a megadott definíció túlságosan *tág*, *vagy* azt kényszerül állítani, hogy az emberen kívül egyetlen más lénynek sincsen tudata, és ezek pusztán materiális testek, vagyis arra az álláspontra kénytelen helyezkedni, mint Descartes. Természetesen a dilemma egyik oldalának elfogadása sem túlságosan kedvező.

Mindazonáltal úgy gondolom, egy vérbeli karteziánus mondhatja erre Strawson védelmében azt, hogy az iménti alternáció második tagja végső soron mégis igaz: a személyeknek *valóban* vannak mentális állapotai, az állatoknak és a számítógépeknek azonban *valójában nincsenek*. A következő módon világíthatja meg karteziánus ellenfelem az álláspontját: Mint az ellenvetés helyesen rámutatott, gyakran mondunk olyanokat, hogy „ez a kutya enni *akar*”, vagy hogy „a Deep Blue meg *akarja* övni a vezérét Kaszparov mesterkedéseivel szemben”. Sőt, ezeknél lényegesen egyszerűbb rendszerek esetében is mondunk ilyeneket, például „ez a fa *érzi*, hogy hamarosan itt a tavasz”⁴, vagy „ez a bimetál *érzi* a hideget és kioldja a biztosítékot”. Ezekben az esetekben azonban a mentális predikátumokat mindössze *metaforikus* vagy *szimbolikus* értelemben használjuk. Az „ez a bimetál *érzi* a hideget és kioldja a biztosítékot” kijelentés esetében mindenki számára teljesen világos, hogy a bimetálnak *nincsenek* (nem lehetnek) mentális állapotai, következésképpen a mondatban szereplő „*érzi*” szó csak is metaforikusan szerepelhet. Éppen így van azonban a fa, a számítógép, vagy éppenséggel a kutya esetében is. Ezekben az esetekben *csak azért* használunk mentális predikátumokat, vagy – Dennett terminológiájával szólva – csak azért tulajdonítunk ezeknek a rendszereknek intencionális állapotokat, hogy *kényelmesen* tudjuk előre jelezni az adott rendszer viselkedésének későbbi állapotait. Amikor például ama feladat előtt állok, hogy megjövendöljem a Kaszparov-Deep Blue mérkőzés során a Deep Blue következő lépését, akkor ugyan módomban áll a számítógép fizikai állapotaiból (a hardveréből) – lévén az determinált rendszer! – kikövetkeztetnem azt, azonban ez az

⁴ Dennett: „Conditions of Personhood” című cikkében megjegyzi, hogy beszélgetett favágókkal, akik egészen „bensőséges” viszonyt építettek ki az egyes fákkal. Ilyen mondatok hangzottak el a favágók szájából: „Ki *akarja* nyújtani az öreg a tagjait, de nem engedem neki”, vagy az egyik napos késő őszi napon: „Azt *hiszi* szegény, hogy itt a tavasz, mekkorát is *fog csalódn*”. Ezek a favágók – meséli Dennett – mindeközben egyetlenegy alkalommal sem az „it”, hanem mindvégig a „he” szóval referáltak a fákra (DENNETT 1978).

eljárás lényegesen körülményesebb és hosszadalmasabb volna, mint ha egyszerűen intencionális rendszernek *tekinténem* a gépet, és azt mondanám, „bizonyára *védeni* fogja a Deep Blue a vezérért ezzel a gyaloggal, tehát a rendszer következő állapota az lesz, hogy a gyalog ide és ide kerül”. Amikor azonban azt mondjuk, hogy „Bertrand Russell megéhezett és most enni *akar*”, akkor ebben az esetben *azon felül*, hogy ezt a kijelentést használhatjuk arra, hogy Bertrand Russell viselkedését (következő állapotát) előre jelezzük, ez a kijelentés arról *is* tudósít bennünket, hogy Russell az „enni akaráss” mentális állapotával rendelkezik. Russell ugyanis – ellentétben a Deep Blue-val – *valóban rendelkezik* mentális állapotokkal. Egyszerűen fogalmazva: míg Russell (és más személyek) esetében a mentális predikátumok tulajdonítása *szó szerint igaz*, addig a személyeknek nem tekinthető rendszerek vagy lények esetében *szó szerint hamis*.

Azt hiszem, hogy a telivér karteziánizmus álláspontja nem utasítható el egy kézlegyintéssel, ugyanis episztemológiai értelemben véve tényleg *nem kizárt*, vagyis tényleg *lehetséges* – ha nem is túlságosan valószínű –, hogy a kutyáknak, macskáknak, lovaknak és malacoknak valójában nincsenek mentális állapotaik, és ráadásul igencsak *valószínű* az, hogy a számítógépek nem is rendelkezhetnek mentális állapotokkal; intuitíve ugyanis nem nagyon tudunk mit kezdeni azzal a feltevéssel, hogy fémdarabokon és szilíciumkristályokon mentális tevékenység realizálódhat. Mindazonáltal úgy gondolom, hogy a telivér karteziánus álláspontja téves.

A karteziánus állásponttal a legfőbb problémám az, hogy a „Bertrand Russell esetében a mentális predikátumok tulajdonítása *szó szerint igaz*” kijelentésben a „szó szerint igaz”, valamint a „Bertrand Russellnek *valóban vannak* mentális állapotai” kijelentésben a „valóban vannak” kifejezéssel – legyen az bármivel, egy fával, egy kutyával vagy egy számítógéppel szemben hangsúlyozva – az égvilágon *semmit nem nyerünk*. Egyszóval: semmi hasznunk nem származik abból, ha így fogalmazunk.

Tételezzük fel, hogy egy személy megtér és vallásos lesz, és ettől a naptól fogva folyton azt hangoztatja, hogy a vallásosság legfontosabb jegye a hitelesség. Ez a hitelesség pedig számára elsősorban azt jelenti, hogy át kell élni a vallás bizonyos misztériumait, e nélkül – mint véli – az egésznek semmi értelme. Nagyon valószínű, hogy ez a személy több alkalommal éles különbséget tesz azok között, akik magukat vallásosnak mondják, de nem élik át a misztériumokat, és azok között, akik átélik. Előfordulhat, hogy ez a személy az egyik nap ekképp fakad ki: „Míg Karcsi és Tomi nem éli át *valóban* a misztériumot, csak azt mondja, hogy át kell élni a misztériumot, addig én *nem csak mondom*, hogy át kell élni a misztériumot, hanem *valóban* át is élem.”

Ha ezt hallanám, azt hiszem, így reagálnék: „Azt a mondatot, amit az imént mondtál, nevezetesen hogy »én *nem csak mondom*, hanem *valóban* át is élem a misztériumot«, *is* lehetséges *csak mondani*.”

Mondok egy másik példát. Tegyük fel, egy számítógép képernyőjén egy alkalommal a következőt olvasom: „Nekem vannak mentális állapotaim.” Természetesen előfordulhat (pontosabban igen valószínű), hogy ennek a számítógépnek nincsenek mentális állapotai, és az iménti szöveg a programozó tréfás kedvének jóvoltából jelent meg a képernyőn. De képzeljük most el azt, hogy egy számítógép monitorán a következő áll: „Míg más számítógépek esetében *csak megjelenik* az a felirat, hogy »nekem vannak mentális állapotaim«, addig az én esetemben nem pusztán erről van szó, nekem ugyanis a többi számítógéppel ellentétben *valóban* vannak mentális állapotaim, és nem csak kiírom a monitoromra a »nekem vannak mentális állapotaim« mondatot.”

Biztosra veszem, hogy egészen hasonló módon reagálnék erre a kiírásra, mint misztikus barátom szavaira: „Az a mondat, ami az imént megjelent ennek a számítógépnek a képernyőjén, nevezetesen, hogy »nekem *valóban* vannak mentális állapotaim, és *nem csak* megjelenik képernyőmön ez a mondat«, is lehetséges, hogy ugyanúgy a vicces-tréfás programozó jóvoltából került a képernyőre, mint az első esetben.”

Mit szeretnék ezzel a két példával mondani? Azt a borzasztóan elemi tényt akarom a karteziánus állásponttal szemben hangsúlyozni, hogy nem artikulálható olyan mondat, amelyet meghallva (vagy egy képernyőn olvasva), legyen az egyes szám első személyben vagy egyes szám harmadik személyben mondva, minden kétséget kizáróan megbizonyosodhatnánk afelől, hogy a beszélőnek vagy a számítógépnek vannak-e mentális állapotai. Egészen egyszerűen azért nem, mert nem képezhető olyan mondat, amelynek pusztá kimondása kizárná a színlelés vagy a hazugság lehetőségét, illetve – a számítógépek esetében – annak a lehetőségét, hogy azt a programozó tréfából előre beprogramozta a rendszerbe. Amennyiben tehát abból indulunk ki, hogy egyedül az ember rendelkezik *valóban* mentális állapotokkal, akkor ezeknek a mentális állapotoknak a létezését a „nekem *valóban* vannak mentális állapotaim, míg a fa, a kutya és a számítógép esetében ezt csak szimbolikusan mondjuk” kijelentésben a „nekem *valóban* vannak” kifejezés hangsúlyozásával nem vagyunk képesek megnyugtató módon igazolni. Szavaink nem alkalmasak arra, pontosabban szólva szavainknak *nem az a funkciója*, hogy egy mentális állapot prezentálódását (például egy misztérium átélését) egyértelműen és kétségbevonhatatlanul artikulálják. Akármilyen trükkös mondatot hallunk is, olyat például, hogy „én tisztában vagyok azzal, hogy mindenki csak mondja, és – bevallom – mind ez ideig én is csak mondtam, de most az egyszer *valóban* megtörtént, hogy átéltem a misztériumot”, ez önmagában, anélkül, hogy *egyéb* körülményeket figyelembe vennénk, nem fog bennünket meggyőzni semmiről. Ebből pedig nyilván egyenesen következik, hogy a „nekem *valóban* vannak mentális állapotaim” kijelentés semmit nem ér, segítségével ugyanis nem világítottuk meg az ember kitüntetett helyét sem az állatokkal, sem a számítógépekkel szemben.

Ellenvethetné valaki azt, hogy érvelésem során egészen egyszerűen csúsztatok, ugyanis összemosisom a „*micsoda* személynek lenni?” kérdést a „*honnnan tudjuk*, hogy egy lény vagy egy gép személy-e?” kérdéssel. Mondhatja valaki azt, hogy az első kérdésre nyugodtan adhatunk választ anélkül, hogy a másodikba belemennénk. Egyszóval, egy *szemantikai* kijelentést (Strawson személy-meghatározását) *episztemológiai* megfontolásokkal bírálom, következésképpen érvelem nem lehet konkluzív.

Ez az ellenvetés jogos, de mégsem véletlen, hogy így jártam el. Tétélezzük fel, hogy Strawson személy-meghatározása alapján véve helyes, és azzal, hogy Strawson azt mondja, „a személy = fizikai és mentális predikátumok logikai alanya”, megadja a „személynek lenni” szükséges és elégséges feltételét. Mindazonáltal egy kicsit (de tényleg csak egy kicsit!) módosítani kell rajta, ugyanis Strawsonnak el kell kerülnie azt, hogy a „személy” fogalmába belevegye a kutyákat, a delfineket és a csimpánzokat, ugyanis ezeket a lényeket *de facto* nem nevezzük személyeknek. Amennyiben Strawson nem akar a telivér karteziánizmus álláspontjára helyezkedni, és azt állítani, hogy *csak* az emberek rendelkeznek mentális állapotokkal, egyedül úgy kerülheti el, hogy személyeknek tekintse a kutyákat, a delfineket és csimpánzokat, ha jelentősen *lecsökkenti* azon mentális predikátumok számát, melyeket a személyeknek nevezett lényekre jogosan predikálhatunk. Bizonyára nem tartozik a kizárólag a személyekről állítható mentális predikátumok közé az „*enni akar*”, vagy a „*hőt érez*”, ugyanis

Strawsonnak el kell ismernie, *szó szerint igaz lehet*, hogy egy kutya, egy delfin vagy egy csimpánz *enni akar* vagy *hőt érzel*. Ezért olyan mentális predikátumokat kell Strawsonnak kiválasztania, amelyeket semmilyen állatra és gépre nem mondunk, például azt, hogy „*szervenélyesen szereti a XIX. századi orosz irodalmat*”, vagy „*észtle-szejtően gyűlöli Heidegger filozófiáját*”. Ebben az esetben Strawson személy-meghatározása a következőképpen alakul át: „*személynek lenni*» = fizikai és *egyed speciális mentális predikátumok logikai alanyának lenni*”.

Ezzel a módosított meghatározással azonban van egy elemi probléma. Az nevezetesen, hogy egyáltalán nem világos, milyen kritériumok szerint teszünk különbséget a speciálisan csak személyekre vonatkozó mentális predikátumok, és az egyéb (maradék) mezei mentális predikátumok között. Pontosabban szólva: amennyiben a mentális predikátumok között ilyen különbséget teszünk, óhatatlanul *episztemológiai* megfontolásokba kell bocsátkoznunk.

Vegyük például a „*szereti a családját*” vagy „*odavan a családjáért*” mentális predikátumot. Nem világos, hogy ezt a mentális predikátumot csak személyekre alkalmazhatjuk-e, vagy jogunkban áll például arra a kutyára alkalmazni, amelyik rendkívül szófogadó, amelyik 100 kilométer távolságból is hazatalált, és amelyik felszínre vonszolta a hó fogságában rekedt gazdáját és családtagjait. Ahhoz, hogy eldöntsük, fel kell-e vennünk a „*szereti a családját*” mint mentális predikátumot a speciálisan csak személyekre vonatkozó mentális predikátumok közé, fel kell tenni azt a kérdést, hogy erre a kutyára vajon *szó szerint igaz-e* a „*szereti a családját*” mentális predikátum, vagy csak szimbolikusan kell értenünk. Ez azonban óhatatlanul ahhoz az *episztemológiai* kérdéshez vezet, hogy „*honnan tudjuk*, hogy ez a kutya rendelkezik-e a »*családja szeretésének*« mentális állapotával”, vagy „*honnan tudjuk*, hogy ez a kutya nem rendelkezik a »*családja szeretésének*« mentális állapotával”. Ha pedig ragaszkodunk ahhoz, hogy illegitim az efféle episztemológiai kérdés Strawson módosított személy-meghatározása kapcsán – lévén az egy *szemantikai* kijelentés –, nem látom, hogyan volnánk képesek meghatározni a speciálisan csak személyekre vonatkozó mentális predikátumok körét *önkényes* eljárás nélkül.

Összefoglalva: azt állítom, hogy a „*P-nek valóban* vannak szellemi állapotai, míg a fáknak, a kutyáknak és számítógépeknek *csak* kényelmességünk okán, pusztán metaforikus értelemben tulajdonítunk szellemi állapotokat” típusú elgondolás semmit nem ér, és a kezdeti dilemma nehézségei, amelybe – mint láttuk – Strawson álláspontja keveredett, teljességgel megmaradtak: Strawson személy-definíciója vagy túlságosan tág, vagy egy terméketlen karteizianizmust eredményez.

MIT JELENT SAJÁTOS (CSAK SZEMÉLYEKRE JELLEMZŐ) TUDATTAL RENDELKEZNI?

Úgy gondolom, Strawson álláspontjának kudarcra egyesek szemében igen jó érvnek tűnhet amellett, hogy a Nagel–Denny-féle elképzelés igaz. Ez az álláspont ugyanis nem vitatja azt, hogy egyes állatok rendelkeznek mentális állapotokkal, hanem mindössze azt szögezi le, hogy mi emberek *sajátos minőségű* (csak ránk, emberekre jellemző) tudatos állapotokkal rendelkezünk: olyanokkal, amelyekkel egyetlen állat és egyetlen számítógép sem rendelkezhet. „*Személynek lenni*” ezek szerint nem más, mint ilyen típusú tudatossággal rendelkezni.

A Nagel–Denny-féle állásponttal szemben azt állítom, hogy a „személy” fogalmának a Nagel-tudat *sem* nem szükséges, *sem* nem elégséges feltétele.

Kezdjük az utóbbival, ezt ugyanis lényegesen egyszerűbb megmutatni. Az állítás tehát a következő: még ha feltételezzük is azt, hogy egy ember rendelkezik Nagel-tudattal, ez önmagában véve *nem elégséges* ahhoz, hogy ezt az embert személynek nevezzük. Képzeld el, hogy egy ember nyitott koponyával fekszik egy műtőasztalon, és az illető – eltérően az agyműtétek során megszokottól – nem éber. Képzeld el továbbá azt is, hogy e műtőasztalon fekvő ember agyát *évek óta* elektromosan stimulálják, és az illető rendelkezik Nagel-tudattal. Mentális színpadán *éppen úgy* megjelenik szülőháza, a gyerekkorában látott rétek, szarvasok stb., mintha csak ott járna és látná. Mindez lehetséges, azonban ezt az embert *nem* neveznénk személynek. Egészen egyszerűen azért nem, mert ez az ember *nem volna autonóm*: Nagel-tudata pusztán agya stimulálásának volna az eredménye. Mivel tehát lehetőségünkben áll elképzelni olyan szituációt, amelyben egy ember rendelkezik ugyan Nagel-tudattal, de nem nevezük személynek, a Nagel-tudat önmagában véve nem elégséges feltétele annak, hogy egy embert személynek mondjunk.

Lényegesen körülményesebb azonban annak a megmutatása, hogy a „személy” fogalmának a Nagel-tudat *nem is szükséges* feltétele. Érvem emellett a következő:

Elképzelek egy olyan világot, ami *pontosan ugyanolyan*, mint ez, amelyben pontosan ugyanúgy történt és történik minden, mint ebben, amelyben pontosan ugyanolyan emberek szerepelnek, mint itt, és ahol ugyanolyan típusú interakciók és szokások vannak érvényben, mint ebben a világban. Ez az általam elképzelt világ tehát *teljesen hasonlít* a mi világunkra. Az egyetlen különbség a mi világunk és az általam elképzelt világ között az, hogy az elképzelt világban egyetlen ember sem rendelkezik Nagel-tudattal. Azt állítom, hogy ebben az elképzelt világban, annak ellenére, hogy egyetlen más ember sem rendelkezik Nagel-tudattal, az embereket mind metafizikai, mind pedig morális értelemben személyeknek nevezném. Ahhoz ugyanis, hogy a körülöttem lévő embereket ne tekintsem személyeknek, egy olyan világot kellene elképzelnem, amelyben minden ember *vagy* automata, *vagy* zombi (azaz olyan lény, amely fizikai tulajdonságait tekintve, valamint funkcionálisan pontosan ugyanolyan, mint mi, csak éppen nem rendelkezik tudattal). Ez utóbbi két esetben azonban egy olyan világot kellene elképzelnem, amely *releváns módon különbözik* a mi világunktól. Egyszóval: ahhoz, hogy egy olyan világot képzeljek el, amelyben a körülöttem levő embereket nem tekintem személynek, *nem elegendő* az, hogy egy olyan világot képzelek el, amelyben az emberek nem rendelkeznek Nagel-tudattal, hanem *az szükséges*, hogy egy olyan világot képzeljek el, ahol az emberek vagy automaták vagy pedig zombik. Érvem tehát arra épül, hogy *míg* lehetőségemben áll elképzelni egy olyan világot, amely pontosan olyan, mint ez a világ, csak éppen a benne szereplő emberek nem rendelkeznek Nagel-tudattal, *addig* nem áll módomban elképzelni egy olyan világot, amely pontosan olyan, mint ez a világ, azonban a benne szereplő emberek vagy automaták, vagy zombik.

Úgy gondolom, ennek az érvnek három előfeltétele van, és hogy az érv kizárólag ezek felől támadható:

Első ellenvetés: Érved azt előfeltételezi, hogy nem tudsz elképzelni egy olyan világot, amely pontosan ugyanolyan, mint ez a világ, de amelyben rajtad kívül minden ember automata. Ez azonban nem igaz, ugyanis elképzelhetnél egy ilyen világot például így: Tételezd fel, hogy létezik egy igen nagy hatalmú diktátor, aki egyben kiváló altató-

orvos is. Tétélezd fel továbbá azt, hogy ez az orvos – különös fondorlatok és cselszövések folytán – világhatalomra tesz szert, és minden egyes embert elaltat, úgy, ahogy ez a komolyabb műtétek során szokás. Tétélezd ugyancsak fel, hogy ez az orvos előállít egy olyan speciális szerkezetet, amelynek segítségével minden egyes mélyaltatásban levő embert éppen úgy képes mozgatni, beszéltetni stb., ahogyan azt a mi világunkban megszoktad. Nos, ez az elképzelt világ *pontosan ugyanolyan* volna, mint ez a világ, azonban a benne szereplő embereket nem mondanád személyeknek, ugyanis ezek a mélyaltatásban levő emberek teljesen automatikusan, a speciális mozgató berendezés jóvoltából viselkednek úgy, ahogy.

Válaszom: Az ellenvetés ott hibázik, hogy valójában az az elképzelt világ, amelyben minden ember mélyaltatásban van, és egy őrült orvos-diktátor mozgatja és beszélteti őket speciális mozgató- és beszéltetőeszköze jóvoltából, legalább egy releváns vonatkozásban *különbözik* a mi világunktól. Bizonyosra veszem, hogy a mi világunkban minden ember *önmozgató*: saját magunk mozgatjuk karunkat, lábunkat, fejünket, hangszálainkat és gégefőnket. Önmozgató képességünknek különféle kémiai, neurofiziológiai, anatómiai stb. folyamatok szolgálnak alapjául; önmozgató képességünk ezekre a folyamatokra való hivatkozással magyarázható. Az ellenvetésben felvázolt lehetséges világban azonban az emberek mozgásának az őrült orvos-diktátor által készített szerkezet szolgál alapjául, következésképpen ahhoz, hogy a mozgásuknak valóságos oksági magyarázatát nyújtsam, olyan erőkre kellene hivatkoznom (ti. az orvos-diktátor mozgatószerkezetéből fakadókra), amelyekről – ha tökéletes az orvos-diktátor trükkje – fogalmam sem lehet. A mi világunkban sikeresen működő fizikai és biológiai magyarázatok ott tehát abszolút csődöt mondanának, hiszen ebben a lehetséges világban az emberek mozgását valójában az orvos-diktátor mozgatószerkezete idézi elő. Következésképpen azt kell, hogy mondjam, az ellenvetésben felvázolt lehetséges világ releváns szempontból *különbözik* a mi világunktól, hiszen a két világban *más és más* az emberek mozgásának a forrása, és a két világban két, egymástól lényegi vonatkozásban eltérő természettudomány nyújtja az emberek mozgásával kapcsolatos események valóságos oksági magyarázatát.

Második ellenvetés: Érved azt előfeltételezi, hogy nem tudsz elképzelni egy olyan világot, amely pontosan ugyanolyan, mint ez a világ, de amelyben minden ember zombi. Ez azonban nem igaz, ugyanis a „zombi” fogalma éppen úgy lett definiálva, hogy fizikai tulajdonságait tekintve, valamint funkcionálisan *pontosan ugyanolyan*, mint mi, csak éppen tudattal nem rendelkezik. Elismerem, az automata-emberek viselkedésének magyarázata esetében egy, a mi világunkban érvényes fizikától különböző fizikát kellene alkalmaznunk, de a zombik viselkedésének magyarázata esetében – mivel azok fizikai tulajdonságaikat tekintve, valamint funkcionálisan megegyeznek velünk – a mi fizikánk teljesen érvényes volna. Nos, a zombi-világ *pontosan ugyanolyan*, mint a mi világunk, azonban a benne szereplő embereket nem neveznéd személyeknek, ugyanis ezek az emberek egyáltalán nem rendelkeznek tudattal.

Válaszom: Először is, egyáltalán nem triviális a zombi-hipotézis, ugyanis egyáltalán nem triviális, hogy *de re* lehetséges olyan világ, amelyben fizikai tulajdonságaikat tekintve, valamint funkcionálisan pontosan ugyanolyan lények élnek, mint mi, azonban nem rendelkeznek tudattal. De ha fel is tesszük, hogy a zombi-hipotézis igaz, abban az esetben is azt kell állítanunk, hogy a zombi-világ és a mi világunk releváns módon *különbözik* egymástól. A két világ fizikai tulajdonságait tekintve pontosan ugyanolyan, mégis a két világ *más és más fizikai történettel* rendelkezik. A zombik ugyanis – *per*

definitionem – mesterségesen előállított lények; úgy jöttek létre, hogy minket, embereket fizikai tulajdonságaink tekintetében duplikáltak, amely eljárás során – hogy úgy mondjam – a tudat vagy a tudatosság nem duplikálódott a fizikai tulajdonságokkal. A *mi* történetünk azonban egészen más. Minket, embereket ugyanis vagy Isten hozott létre egy *ex nihilo* teremtő aktussal, vagy az evolúció során jöttünk létre. Azaz abból, hogy két világ fizikai tulajdonságait tekintve pontosan ugyanolyan, *nem következik*, hogy a két világ ugyanolyan, ugyanis a két világ története lehet éppenséggel teljesen különböző.

Harmadik ellenvetés: Tétélezzük fel, hogy az előző két ellenvetésre adott válaszod helyes, és igaz az, hogy nem tudunk elképzelni olyan világot, amely pontosan olyan, mint a *mi* világunk, de amelyben minden ember vagy automata, vagy zombi. Mégsem tetted ezzel világossá azt, hogy abban a világban, amelyben egyetlen ember sem rendelkezik Nagel-tudattal, az embereket jogosan nevezheted személyeknek. Nagel-tudattal nem rendelkezni ugyanis nem más, mint zombinak lenni, és a zombikat – ahogy magad is mondod – nem nevezzük személyeknek. Egyszerűen szólva: milyen alapon nevezel személynek olyan lényeket, amelyek nem rendelkeznek tudattal?

Válaszom: Érvelésem során nem azt állítottam, hogy egy tudattal nem rendelkező lényt személynek neveznék, hanem azt, hogy el tudok képzelni olyan világot, amelyben az emberek nem rendelkeznek Nagel-tudattal, de akiket ettől függetlenül személynek neveznék. Egyszóval: különbséget lehet tenni a „Nagel-tudattal rendelkezés” és a „tudattal rendelkezés” fogalma között. A „Nagel-tudattal rendelkezés” fogalma szerves kapcsolatban van a zombi-hipotézissel, ugyanis – mivel a Nagel-tudat fizikalisztikusan nem írható le és nem magyarázható – egy, a *mi* világunkban élő ember e hipotézis szerint *nem más*, mint zombi + Nagel-tudat. Precízebben mondva: egy, a *mi* világunkban élő ember nem más, mint egy fizikai tulajdonságait tekintve és funkcionálisan pontosan ugyanolyan organizmus és (ezen felül) Nagel-tudat. Ezt a hipotézist azonban egyáltalán nem kötelező elfogadni. Hiszen elképzelhetek olyan világot, amelyben egyetlen ember sem rendelkezik Nagel-tudattal, e világban szereplő embereknek a tudata ugyanis fizikai entitásokon *realizálódik* és *nem emergálódik*. Pusztán arra való hivatkozással, hogy ebben a világban az emberek nem rendelkeznek Nagel-tudattal, vagyis olyan tudattal, amely nem magyarázható fizikalisztikusan, *nem fogom azt állítani*, hogy ezek az emberek nem személyek. Ahhoz ugyanis, hogy egy lehetséges világban az emberekről azt állítsam, hogy nem személyek, ennél *többre* van szükség. Vagy arra, hogy ebben a világban az emberek automaták legyenek, vagy arra, hogy zombik: olyan lények, amelyek nem Nagel-tudattal nem rendelkeznek, hanem semmilyen tudattal.

MIT JELENT SZABAD AKARATTAL RENDELKEZNI?

Idáig azt fejtettem ki, hogy más dolog „személynek lenni”, és más dolog „Nagel-tudattal rendelkezni”, ugyanis ez utóbbi a „személy” fogalmának sem nem szükséges, sem nem elégséges feltétele. Amikor az előző rész elején amellet érveltem, hogy a Nagel-tudat nem elégséges feltétele a „személy” fogalmának – mivel az évek óta nyitott koponyával fekvő ember Nagel-tudata pusztán agya stimulálásának az eredménye –, úgy fogalmaztam, hogy ez az illető azért nem személy, mert nem autonóm. Így aztán, ahogy Strawson kudarca átvezetett bennünket Nagel és Denny álláspontjáig, úgy

most Nagel és Denny kudarca vezet át bennünket Frankfurt álláspontjáiig, amely szerint a „mit jelent személynek lenni?” kérdésre az az egyedül üdvözítő felelet adódik, hogy „»személynek lenni« nem más, mint *szabad akarattal* (vagyis *autonomiással*) rendelkezni”. Azt a lényt nevezzük tehát Frankfurt szerint személynek, aki képes *másodfokú* vágyak vagy akarások kialakítására, aki *autonóm* abban a tekintetben, hogy milyen akarral rendelkezik.

Frankfurt javaslatával szemben azt állítom, hogy a „szabad akarat” fogalmának nem adható megnyugtató filozófiai analízise; akárhogyan csűrjük-csavarjuk is, a „szabad akarat” fogalma *inkonzisztens* jelleget mutat. A következő megfontolásokat ajánlom⁵:

A *determinizmus* tétele azt mondja ki, hogy minden egyes E esemény elkerülhetetlen következménye az azt megelőző eseményeknek. Más szóval azt, hogy amennyiben ilyen és ilyen események bekövetkeztek, akkor E-nek be *kell* következnie, azaz nem fordulhat elő (nem lehetséges), hogy E ne következzen be.

Az „emberi szabadság” fogalma kapcsán szokás megkülönböztetni a *cselekvés* és az *akarat* szabadságát. A cselekvések szintjén megfogalmazva a szabadság tézise ekképp hangzik:

Egy adott P személy abban az esetben cselekszi A-t szabadon, amennyiben *mind A* megtétele, *mind pedig A* nem megtétele *lehetséges* számára, és *egyedül P-n múlik* az, hogy melyiket cselekszi.

Múlt időben megfogalmazva:

Egy adott P személy abban az esetben cselekedte A-t szabadon, amennyiben P-nek A cselekvés időpontjában *lehetőségében állt* nem-A-t cselekedni, vagy – más szavakkal – amennyiben P A cselekvése időpontjában *képes lett volna* nem-A-t cselekedni, és *egyedül P-n múlt*, hogy A-t cselekedte.

Az akarat szintjén megfogalmazva a szabadság tézise így hangzik (és ez az, amit Frankfurt állít):

Egy adott P személy V akarata akkor szabad, ha P-nek *lehetőségében áll*, hogy más akarattal rendelkezzen, mint V, és *egyedül P-n múlik* az, hogy V és nem nem-V akarral rendelkezik.

Ugyanez múlt időben:

Egy adott P személy V akarata akkor volt szabad, ha P-nek *lehetőségében állt*, hogy V akarásának időpontjában más akarral rendelkezzen, mint V, és *egyedül P-n múlt*, hogy V és nem nem-V akarral rendelkezett.

A determinizmus *versus* szabad akarat kapcsán felmerülő probléma egészen pontosan a következő: A determinizmus szerint minden esemény (így minden egyes akarás is!) olyan, aminek meg *kell* vagy meg *kell* történnie, a szabad akarat hívei szerint azonban egy személynek egyes esetekben *lehetőségében áll* megtenni vagy nem megtenni valamit, illetve *lehetőségében áll*, hogy V vagy nem-V akarral rendelkezzen. Míg tehát a szabad akarat hívei *legalább két* lehetőségre formálnak igényt, addig a determinizmus álláspontján állók *mindössze egyet* engedélyeznek.

A determinizmus *versus* szabad akarat probléma egyik, talán legtöbbek által osztott megoldási, vagy még inkább feloldási kísérlete az ún. *kompatibilista* stratégia. E szerint valójában nem áll fenn konfliktus a szabad akarat és a determinizmus között, ugyanis a két álláspont *két különböző értelemben vett* lehetőség-fogalomra apellál. Amikor az emberi szabadság híve azt mondja, hogy „P akkor cselekedte A-t szaba-

⁵ Túlzás nélkül: e rész gondolatmenete nem jöhetett volna létre Farkas Katalin segítségével.

don, ha *lehetőségében állt* nem-A-t is cselekedni”, akkor a „lehetőségében állt” kifejezés hangsúlyozása *nem áll szemben* a determinizmus tézisével.

A három legismertebb kompatibilista javaslat a következő:

Első javaslat: Szabadon cselekedni nem egyéb, mint *külső korlátozások nélkül* cselekedni. Márpedig minden további nélkül lehetséges az, hogy egy cselekedet külső körülmények által *nem korlátozott*, azonban – mondjuk – különféle neurofiziológiai összetevők oksági hatása folytán *determinált*.

Válaszom: Véleményem szerint a kompatibilizmus ezen megfogalmazása abszolút célt téveszt, ugyanis végső soron nem mond többet annál, mint hogy a legtöbb cselekvésünk során semmiféle kényszert nem érzünk. Ez azonban a szabadságnak csak igen-igen szűk, és nem túlságosan sokunkat kielégítő értelmezése. Ez a nézet ugyanis – úgy tűnik – elvi okok folytán *kizárja* a „felelős személy” fogalmát, hiszen e nézet szerint nem teljesül a szabadságnak az a feltétele, amit fentebb úgy fejeztem ki, hogy „*egyedül P-n* (mint felelős ágensen!) *műlik* az, hogy A-t vagy nem-A-t cselekszi”. A kompatibilizmus ezen javaslata szerint P neurofiziológiai állapotai *következtében* cselekszi azt, amit cselekszik: P ugyan nem érez kényszert – hiszen az ember neurofiziológiai állapotai nem kényszerként jelentkeznek –, valójában azonban P *nem tehet másként*, mint amire neurofiziológiai állapota okságilag determinálja. Ez a következmény azonban *összeegyeztethetetlenül* szemben áll az „emberi szabadságból fakadó felelőség”, vagy az „emberi autonómítás” fogalmával.

Második javaslat: Az a kijelentés, hogy „P cselekedhetett volna másképp”, ekvivalens azzal a kijelentéssel, hogy „*amennyiben P másképp választott volna*, abban az esetben P *másképp cselekedett volna*”. Teljesen nyilvánvaló az, hogy ez utóbbi kijelentés *összeegyeztethető* a determinizmussal, hiszen még abban az esetben is, ha egy ember valamennyi cselekedete kauzálisan determinált, értelmesen (és igaz módon) állítható, hogy *ha* az adott személy másképp választott volna, *akkor* másképp cselekedett volna. Más szóval, a szabad cselekedetnek *mindössze* az az előfeltétele, hogy az illető másképp is tehetett volna, értve ez alatt azt, hogy másképp cselekedett volna, *ha* a világ a cselekvés előtt másmilyen.

Válaszom: Ez a javaslat véleményem szerint megint csak célt téveszt, mégpedig azért, mert a két kijelentés minden látszat ellenére *mégsem ekvivalens*: más és más igazságértékkel rendelkezik. Lehetséges, hogy míg az utóbbi kijelentés igaz, addig az első hamis. Mondhatjuk ugyanis azt, hogy „*ha* P másként döntött volna, *akkor* másképp cselekedett volna, valójában azonban P *mégsem tudott másképp cselekedni*, ugyanis *nem állt módjában másképp dönteni*”. A kompatibilizmus tehát csak abban az esetben volna elfogadható javaslat, amennyiben azt *is* igazolni tudná, hogy „P-nek lehetőségében állt másképp *dönteni*, hogy aztán másképp cselekedjen”. A determinizmus azonban *összeegyeztethetetlen* ezzel, hiszen egy döntés is egy esemény, és a determinizmus szerint minden egyes esemény elkerülhetetlen következménye az őt megelőző eseményeknek.

Harmadik javaslat: Induljunk ki abból a Ryle és Wittgenstein által képviselt álláspontból, miszerint gyökeresen különböznek egymástól az *indokok* és az *okok*. E szerint a fiziológia elvileg lehet determinisztikus, azonban a testi mozgások *oksági* (fiziológiai) magyarázata *különbözik* az emberi cselekvések *indokokon alapuló* magyarázatától. Következésképpen az emberi cselekvések determinista magyarázatának feltételezése egészen egyszerűen *konceptuális hiba*.

Válaszom: Tegyük fel, hogy a determinista (okszági) magyarázatok a fizikai világra korlátozottak. Ha a determinizmus igaz, és ennek értelmében egy adott időpillanatban fizikailag determinált az, hogy nem emelkedhet fel a karom, akkor nekem nem áll módomban felemelni a karomat, akármilyen indokaim is vannak arra nézve, hogy felemeljem. Egyáltalán nem világos tehát az, hogy hogyan lehetünk képesek olyasmire, ami okszájilag lehetetlen.

A kompatibilizmus álláspontjával szemben az előzőeken felül – véleményem szerint – egy kikezdzhetetlen ellenvetés is felhozható. Az érv Inwagen-tól származik (INWAGEN 1975), és a következőképpen foglalható össze:

(P1) Senkinek nincsen lehetősége arra, hogy igazzá tegyen egy olyan állítást, amely ellentmond a természettörvényeknek.

(P2) Senkinek nincsen lehetősége arra, hogy megváltoztassa a múltat, vagyis hogy két ellentmondó múltról szóló állítás közül a hamisat igazzá tegye.

(P3) (P2-ből) Egy S személynek csak abban az esetben van lehetősége egy adott időpontban végrehajtani egy A cselekvést, ha egy teljes mértékben a múltról szóló p igazság esetében el tudja érni, hogy a $\neg p$ és S végrehajtja A-t⁶ esete fennálljon.

(K) (P1)-ből és (P3)-ból, valamint a determinizmus téziséből együttesen az következik, hogy senkinek sincsen módja azt tenni, amit ténylegesen tesz.

Mondom egyszerűbb formában: Legyen (P0) annak a világnak a teljes leírása, amelyben még nem volt ember. Legyen (L) a természeti törvények összessége, és legyen (P) egy tetszőleges igaz proposíció. A determinizmus tétele azt mondja ki, hogy (P0) és (L) együttesen implikálja (P)-t. No mármost, (P0)-t és (L)-t senki nem választotta, továbbá (P0)-t és (L)-t senki semmilyen módon nem befolyásolhatja. Vagyis, ha a determinizmus tétele igaz, akkor ebből az az abszurditás következik, hogy egyetlen igaz proposíció sem múltott vagy múlik senkin. Az a tény, hogy itt ülök, és írom ezt a tanulmányt, valójában nem rajtam múlik: a világ teljes leírása, valamint a természettörvények együttesen határozzák meg, olyan „hatalmak” tehát, amelyekre semmiféle befolyásom nem lehet. Nagyon egyszerűen szólva: ahhoz, hogy képesek legyünk mást tenni, mint amit ténylegesen teszünk, vagy a természettörvényeket kellene megváltoztatnunk, vagy pedig a múltat, de természetesen ezek közül egyik sem áll hatalmunkban. Ebből következően, ha a determinizmus igaz, akkor nem lehetséges szabad akarat, a kettő szükségképpen összeegyeztethetetlen.

Azt hiszem, ezek után nem tévedek nagyot, ha azt állítom: annak, aki fenn akarja tartani a „szabad akarat” fogalmát, azt kell állítania, hogy *létezik* szabad akarat, és a determinizmus tézise *hamis*. E nézet ellen azonban – bármilyen különösen hangzik is! – megint csak egy végzetes ellenvetés szól:⁶

Tételezzük fel, hogy létezik egy Ikerföld, ami *pontosan ugyanolyan*, mint a mi Földünk. Ezen az Ikerföldön minden valaha élt és ma élő földi embernek megvolt és megvan a maga hasonmása, és e hasonmások története, mind fizikai, mind pedig mentális vonatkozásban, pontról pontra megegyezik a földi emberek történetével. Egy szóval, az Ikerföldön minden egyes valaha élt és ma élő földi ember minden eddigi cselekedetének, gondolatának, érzésének, interakciójának stb. megvolt és megvan a maga pontos megfelelője. Természetesen nekem is megvan a magam pontos ikerföldi hasonmásom. Tegyük fel azonban, hogy hetven év abszolút megegyező életút után

⁶ Az érvet R. Kane-től kölcsönöztem (KANE 1996).

egy ponttól kezdve életutunk hirtelen eltér. Mindketten egy útkereszteződéshez érünk, én itt a Földön jobbra fordulok, hasonmásom azonban ott az Ikerföldön balra fordul.

Mit akar ez a gondolkíséret mondani? A következő problémára hívja fel a figyelmet: mint láttuk, szabad cselekvésemnek vagy szabad akaratomnak az a feltétele, hogy mindaz, ami a múltban történt, *ne* határozza meg a cselekedetemet, illetve az akaratomat. Felmerül azonban a kérdés: amennyiben nem az előzmények, akkor pontosan *mi* határozza meg egy adott cselekedetemet, illetve egy adott akarásomat? Térjünk vissza egy pillanatra az iménti gondolkísérlethez! Elképzelem, hogy a szétválás utáni pillanatban vagyok, és míg hasonmásom az ikerföldi útkereszteződésben balra gyalogol, addig én a földi útkereszteződésben jobbra haladok. Képzeletemben visszafordítom az idő kerekét, és lassan közelítek ahhoz az időpillanathoz, amikor elváltak útjaink. Mi történik ebben a pillanatban? Hogyan vagyok képes megmagyarázni hirtelen elválásunkat?

Egyedül a *véletlenre* való hivatkozással. *Véletlenül* valami más történt ikerföldi hasonmásom elméjében, mint az én elmémben. Teljesen véletlenül, azaz *mindenféle előzmény nélkül*, elmémbe hasított, hogy „fordulj jobbra”, míg ikerföldi hasonmásomnak ugyancsak teljesen véletlenül elméjébe hasított, hogy „fordulj balra”? Vagy, teljesen véletlenül, ikerföldi párom agyában egy elektron felfelé perdült, az én agyamban viszont a megfelelő elektron lefelé? Az ilyen típusú válaszok nem különösebben heurisztikusak annak ismeretében, hogy hetven évig elménkben és agyunkban abszolút izomorf folyamatok mentek végbe.

A tanulság a következő: *vagy* azt kényszerülünk állítani, hogy minden egyes cselekedetünk vagy akarásunk *determináltan* jön létre, vagyis azt, hogy minden egyes cselekedetünket vagy akarásunkat az azt megelőző események határozzák meg, *vagy* azt kényszerülünk állítani, hogy egyes cselekedeteink vagy akarásaink *véletlenül*, azaz mindenféle előzmény nélkül, vagy éppenséggel annak ellenében jönnek létre. Az elszomorító az, hogy akármelyik álláspontot tesszük is magunkévá, azzal a „szabadság” vagy „szabad akarat” fogalma összeegyeztethetetlen lesz. Sem a determinizmussal, sem pedig a „véletlen” fogalmán alapuló indeterminizmussal nem egyeztethető össze ugyanis a szabadság. Harmadik lehetőség pedig nincs. Következésképpen a „szabadság” *inkonzisztens, elemezhetetlen* fogalom. Következésképpen *nem tudjuk*, mit jelent autonóm személynek, autonóm ágensnek lenni.

VÉGSŐ KONKLÚZIÓK

Ebben a tanulmányban azt igyekeztem megmutatni, hogy a „személy” metafizikai fogalmának három szükséges feltétele közül egyiknek sem vagyunk képesek megnyugtató filozófiai analízisét nyújtani. Más szóval azt, hogy Strawson, Denny és Frankfurt személy-meghatározásai csődöt mondanak. Azt hiszem, sikerült megmutatnom, hogy (1) nem tudjuk azt, mi a különbség a mentális predikátumok szimbolikus és tényleges tulajdonítása között; hogy (2) a „személy” fogalmának nem a „Nagel-tudat” fogalmára kell alapozódnia; és hogy (3) a „szabadság” fogalma mind a determinizmussal, mind a véletlenre alapozó indeterminizmussal összeegyeztethetetlen, és nincsen harmadik lehetőség. Ezekből a megfontolásokból számomra az következik, hogy *nem tudjuk*, mit jelent személynek lenni.

Ez természetesen abszolút negatív és kissé talán kiábrándító eredmény. Felmerülhet tehát a kérdés: létezik-e kiút? Nem lehetséges-e az, hogy a „személy” fogalmát egészen másként, más modorban kell tárgyalni, mint ahogy ezt én tettem? Talán igen. A próba kedvéért térjünk vissza az utolsó rész végéhez, amikor is azt állítottam, hogy a „szabadság” fogalma sem a determinizmussal, sem pedig az indeterminizmussal nem egyeztethető össze, és harmadik lehetőség nincs. Ebből én *akkor* arra következtettem, hogy a „szabadság” inkonzisztens fogalom, ami azt jelenti, hogy nem tudjuk, mit jelent autonóm (szabad) személynek, autonóm (szabad) ágensnek lenni. Következtessünk most másként. Ha igaz az, hogy létezik autonóm személy (és azt kell állítanunk, hogy létezik, hacsak nem kívánjuk tagadni az „emberi felelősség” fogalmát!), és ha ugyancsak igaz az, hogy az „autonóm személy” fogalma nem egyeztethető össze sem a determinizmussal, sem pedig az indeterminizmussal, *akkor nem lehet másként*, mint hogy léteznek olyan események, amelyek egyrészt nem véletlenül következnek be, másrészt nem is más események okozataként, hanem egyedül az Én mint tovább redukálhatatlan instancia által. Egyszerűen szólva, *azt kell állítanunk*, hogy léteznek olyan események, amelyeknek nem más események az okai, hanem maga az ember, maga az Én mint ágens.

Ez az álláspont Chisholmtól származik⁷, aki a következőképpen képzei el az *ágens-okozást*: Vegyük azt az egyszerű esetet, hogy egy ember egy bottal arrébb lök egy kavicsot. Chisholm szerint ezt a jelenséget úgy kell magyarázni, hogy a bot mozgását az illető karjának mozgása okozza, az illető karjának mozgását különféle izmok összehúzó mozgása okozza, az izmok mozgását bizonyos agyban végbemenő események okozzák, *de* az illető agyában bekövetkezett eseményeket immár az Én mint ágens okozza. Az Én és az Én agyállapotai között tehát *ágens-okozás* (vagy *immanens-okozás*) van, vagyis az illető izmainak mozgását előidéző, az agyban végbemenő eseményeket nem más, az agyban végbemenő esemény vagy események okozzák, hanem a cselekvő Én. Vagy, ahogy ezt Chisholm kissé patetikus modorban kifejezi: „Aki szabadon cselekszik, annak olyan képessége van, mint amit néha istennek tulajdonítottak: mindenki közülünk, aki csak cselekszik – mozdulatlan mozgató. Cselekvésünk során *megettörtétté teszünk* (we made something happen) bizonyos eseményeket, noha valójában semmilyen esemény nem okozza ezen események bekövetkeztét.” (CHISHOLM 1966, 23.)

Ha elfogadjuk Chisholm javaslatát, könnyedén feloldhatjuk az előző fejezet csüggesztő konklúzióját, nevezetesen azt, hogy egy esemény *vagy* determináltan, *vagy* pedig a puszta véletlen szeszélyéből következik be. Ekképp: a „véletlen” *kétféle*t jelenthet. Ha azt jelenti, hogy „előzmények által nem determinált”, akkor a „vagy determinizmus, vagy véletlen” elképzelés hívének külön be kell bizonyítania azt, hogy az ilyen értelemben vett véletlen cselekvések kívül esnek az ágens tudatos kontrollján, ellenkező esetben ugyanis Chisholm mondhatja azt, hogy ágens-okozta eseményről van szó. Ha viszont ezenfelül még a cselekvő szubjektum passzivitását is implikálja, akkor a „vagy determinizmus, vagy véletlen” elképzelés híve egészen egyszerűen feltételezi a bizonyítandót, ugyanis Chisholm tagadja azt, hogy a cselekvő szubjektum passzív volna választásai tekintetében.

⁷ Chisholm az álláspontját többek között „Human Freedom and the Self” és „Freedom and Action” című tanulmányában fejti ki (CHISHOLM 1982; 1966).

Elfogadjuk Chisholm javaslatát, vagy utasítsuk el? Chisholm érve *egyrésztől* egy jogosnak tűnő transzcendentális érv, *másrésztől* azonban az általa bevezetett „ágens-okozás” fogalma teljességgel a misztikus homályba vész. Melyik evidenciára hallgassunk? Arra-e, amire Chisholm hivatkozik, nevezetesen hogy léteznek felelős emberi személyek (vagy még inkább létezniük kell), és ez csak úgy lehetséges, ha azt állítjuk, hogy létezik ágens-okozás, vagy inkább arra, hogy hiába tesszük fel (vagy kell feltennünk), hogy létezik ágens-okozás, ha az égvilágon semmit nem tudunk róla, és ráadásul természettudományos világképünkkel is abszolút mértékben szemben áll?

Erre a dilemmára nem tudom a választ, és azt sem tudom, egyáltalán hogyan kellene választ találnom rá. Talán az egész tanulmányomból az a tanulság, hogy *nem csak* azt nem tudjuk, mit jelent személynek lenni, hanem azt *sem*, hogy hogyan kellene megkísérelnünk a válaszadást.

IRODALOM

- CHISHOLM, Roderick M. 1966. Freedom and Action. In Lehrer, Keith (ed.): *Freedom and Determinism*. New York, Random House.
- CHISHOLM, Roderick M. 1982. Human Freedom and the Self. In Watson, Gary (ed.): *Free Will*. Oxford, Oxford University Press.
- DENNETT, Daniel C. 1978. Conditions of Personhood. In *Brainstorms*. Hassocks, Harvester Press.
- DENNY, M. 1986. *The Person and the Nagel-Consciousness*. Philadelphia, Temple University Press.
- FRANKFURT, Harry G. 1971. Freedom of the Will and the Concept of a Person. *Journal of Philosophy*, 68.
- INWAGEN, Peter van 1975. The Incompatibility of Free Will and Determinism. *Philosophical Studies*, 27.
- KANE, Robert 1996. *The Significance of Free Will*. New York, Oxford University Press.
- NAGEL, Thomas 1979. What is it Like to be a Bat? In *Mortal Questions*. Cambridge, Cambridge University Press.
- RORTY, Amélie 1973. Persons, Policies and Bodies. *International Philosophical Quarterly*, 13.
- SAPONTZIS, S. F. 1981. A Critique of Personhood. *Ethics*, 91.
- SCHLICK, Moritz 1949. Meaning and Verification. In Feigl, Herbert – Sellars, Wilfrid (eds.): *Readings in Philosophical Analysis*. New York, Appleton–Century–Crofts.
- STRAWSON, P. F. 1962. Persons. In Chappell, Vere C. (ed.): *The Philosophy of Mind*. Englewood Cliffs, NJ, Prentice Hall.